Structure determination from screw-disordered fibres

R. P. Millane* and J. L. Eads

Whistler Center for Carbohydrate Research and Computational Science and Engineering Program, Purdue University, West Lafayette, Indiana 47907-1160 USA

* Current address: Department of Electrical and Computer Engineering, University of Canterbury, Christchurch, New Zealand

Received 14th March 2002; accepted 16th April 2002.

Fibre diffraction analysis has traditionally used x-ray data from either noncrystalline or polycrystalline fibres. In some cases however, only data from specimens with intermediate ordering are available. A simple method for incorporating the effects of screw-disorder into structure determination, using a conventional refinement program, is described. The method is applied to diffraction data from a screw-disordered polynucleotide fibre.

Introduction

The degree of order in oriented specimens used for x-ray fibre diffraction analysis varies greatly. In a noncrystalline fibre, the diffracting particles are randomly rotated about the axis of orientation, and the diffraction pattern shows continuous intensity on layer lines. In a polycrystalline fibre, the molecules form small, well-ordered crystallites that are randomly rotated, and the diffraction pattern consists of discrete Bragg reflections. It is therefore straightforward to calculate the intensity diffracted by models of noncrystalline and polycrystalline fibre specimens, and for this reason, almost all polymer structures so far determined by x-ray fibre diffraction analysis have utilized one of these two kinds of specimen [1,2].

It is not uncommon, however, for fibres to give diffraction patterns that display both sharp reflections and continuous intensity on layer lines [3-6], indicating that the packing of the molecules is neither ideally noncrystalline nor ideally polycrystalline. In fact, a spectrum of such disordered fibres has been observed [3,4]. The Bragg reflections are often confined to the centre of the diffraction pattern, giving way to continuous intensities at high resolution, indicating that the packing of the molecules in the crystallites is disordered in some way. The diffraction pattern recorded from such a specimen depends not only on the molecular and crystal structures, but also on the type and degree of disorder in the specimen. If the disorder is *uncorrelated*, then the diffraction can be separated into the sum of continuous and Bragg components. These components have the same general character as those from noncrystalline and polycrystalline specimens, but are different in detail.

Disordered fibres have generally not been used for structure determination because of the difficulty of calculating the diffraction from such specimens and of refinement. In cases where they have, an approximate method has been used in which molecular and crystal structures are co-refined against the continuous and Bragg diffraction, treating the continuous intensity data as if they are from a noncrystalline fibre, and the low resolution Bragg data as if they are from an ideal polycrystalline fibre [7,8]. Such an analysis is only approximately valid however, since it ignores the effects of the disorder on the diffracted intensities. Stroud and Millane [5,6] have recently reported a comprehensive analysis of diffraction by fibres that contain various kinds of packing disorder, and use of this theory allows the possibility of determining structures from these kinds of specimens while rigorously incorporating the effects of the disorder. We describe here a method for implementing such an approach and its application to diffraction data from a screw-disordered polynucleotide fibre.

There are two reasons for pursuing structural analysis of molecules in disordered fibres. First, some biopolymers pack only as disordered fibres, so that this is the only route by which structural information can be obtained. Second, since the disorder produces an averaging of the effects of specific intermolecular interactions (that are present in a polycrystalline fibre), structures in disordered fibres are more likely to represent intrinsic molecular structures free of these intermolecular effects.

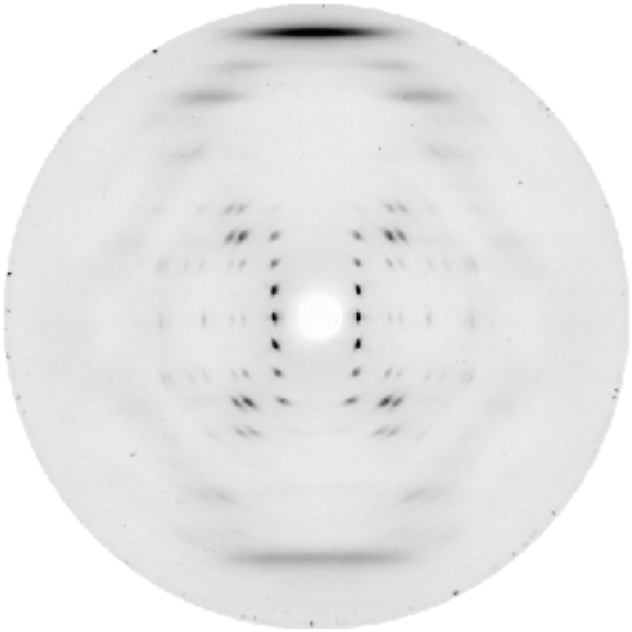

Figure 1: Fiber diffraction pattern from a-poly(dA)·poly(dT) (from Ref. 11)

Diffraction by fibres

The intensity diffracted by a noncrystalline fibre at cylindrical polar radius R on layer line l is given by

$$I_l(R) = \sum_n |G_{nl}(R)|^2 \qquad (1)$$

where the $G_{nl}(R)$ are the Fourier-Bessel structure factors and the sum is over the integers that satisfy the helix selection rule. In structure determination from such a specimen, molecular models are refined against the data $I_l(R)$. For a polycrystalline fibre, assuming a monoclinic unit cell with c parallel to the rotation axis, the intensity of the i-th spot on the l-th layer line, I_{il}, is given by

$$I_{il} = \sum_{(h_i,k_i)} |F_{h_i k_i l}|^2 \qquad (2)$$

where the F_{hkl} are the usual crystallographic structure factors, and the sum is over the reciprocal lattice points $(h_i k_i l)$ that overlap in the spot (i,l). The structure factors can be calculated in the usual way, or in terms of the Fourier-Bessel structure factors as

$$I_{il} = \sum_{(h_i,k_i)} \left| \sum_n G_{nl}(R) \exp(in(\psi_{h_i k_i} + \pi/2)) \right|^2 \qquad (3)$$

where $(R_i, \psi_{hiki}, l/c)$ are the cylindrical polar coordinates of the reciprocal lattice point $(h_i k_i l)$. In structure determination from such specimens, molecular and crystal structures are refined against the data I_{il}.

The packing disorder in the microcrystallites of disordered polycrystalline fibres is conveniently described in terms of *lattice* disorder and *substitution* disorder [5]. Lattice disorder is due only to deviations in the positions of the molecules away from their positions on a regular undistorted lattice. Substitution disorder is due to variations in the orientation of the molecule (or in the kind of molecule) at each site. If the lattice and substitution disorders are independent, and the distortions at different lattice points are uncorrelated, then the diffraction can be separated into the sum of continuous and Bragg components [5]. If the Cartesian components of the lattice disorder are independent, and the x and y components have equal variances, then general expressions have been derived for the continuous and Bragg diffraction in terms of the statistics of the disorder [5].

We consider here the case of screw disorder (a form of substitution disorder) that is quite common with molecules of high helix symmetry (such as polynucleotides). Screw disorder occurs when the helical molecules in a crystallite randomly "screw in and out" of a plane normal to the rotation axis, as a result of rigid body motion constrained by the interlocking of grooves and protuberances of neighbouring molecules. If the screw disorder is completely random, its symmetry is the same as the molecular helix symmetry (as would usually be expected), and the molecular helix is integral (one helix turn in one c-repeat), then the continuous and Bragg components of the diffraction are given by [4,5]

$$I_l^{(sd)}(R) = \sum_n |G_{nl}(R)|^2 - |G_{ll}(R)|^2 w_{lattice}(R, l/c) \qquad (4)$$

and

$$I_{il}^{(sd)} = N_{il} |G_{ll}(R)|^2 w_{lattice}(R, l/c) \qquad (5)$$

The *lattice disorder weight* $w_{lattice}(R,Z)$ is given by

$$_{ice}(R,Z) = \exp(-4\pi^2(R^2\sigma_{lat}^2 + Z^2\sigma_{axial}^2)) \qquad (6)$$

where σ_{lat}^2 and σ_{axial}^2 are the variances of the distortions of the lattice sites normal (or "lateral") and parallel (or "axial") to the direction of orientation, respectively, N_{il} is the number of overlapping reflections in spot *(i,l)*, and the superscript (*sd*) denotes "screw-disordered." (Note that the term $Z^2\sigma^2_{axial}$ is missing in equations (75) and (76) of Ref. 5).

Structure determination incorporating the effects of disorder

Structure determination using data from disordered fibres in general requires optimisation of molecular and disorder parameters to obtain the best match between the calculated and measured diffraction. For the case of screw disorder however, the simplicity of the expressions (4) and (5) allows a simple refinement protocol to be developed based on use of a conventional fibre refinement program such as LALS [9], that rigorously incorporates the effects of the disorder. This is the approach we have taken. In many cases, most of the available data (and all the high resolution data) are continuous, and the

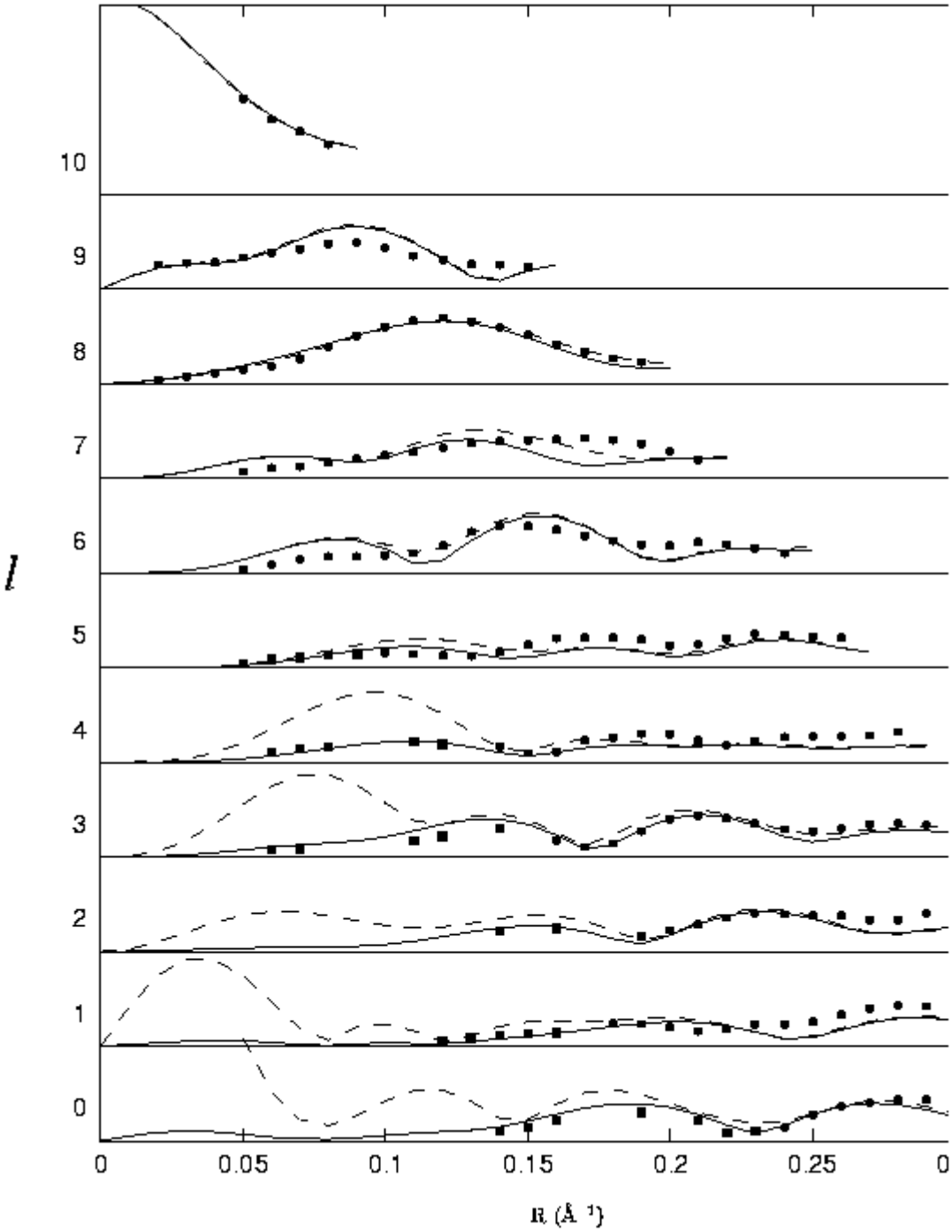

Figure 2: Observed x-ray amplitudes in the resolution range ~6-3Å (circles), and in the resolution range ~8-6 Å (squares). Amplitudes calculated from refined models based on screw-disordered (---) and noncrystalline (- - -) fibres.

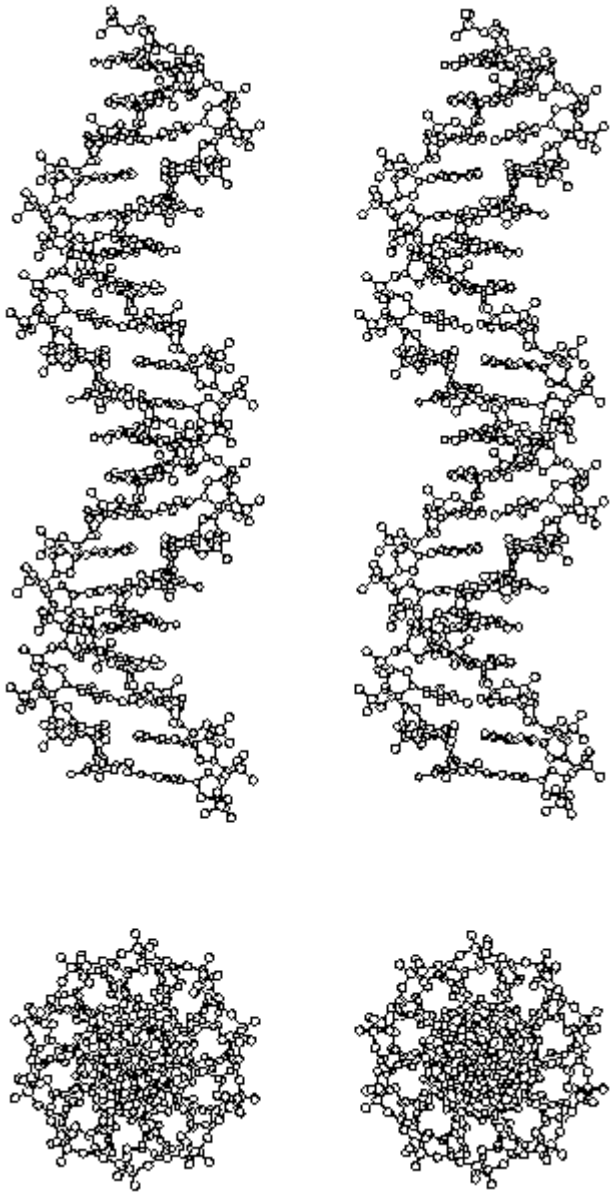

Figure 3: Structures of a-poly(dA)·poly(dT) viewed normal and parallel to the helix axis, determined as described here incorporating the effects of the screw disorder (left), and as determined in Ref. 8 (right).

refinement scheme we describe is based on using the continuous data. The approach is to iteratively modify the data to "take out" the effect of the disorder, allowing refinement based on a noncrystalline fibre. This is done as follows.

1. Obtain an initial estimate of the structure using conventional methods based on the continuous data and assuming a noncrystalline specimen.

2. Select values of σ_{lat} and σ_{axial} .

3. Using the current estimate of the molecular structure, calculate the continuous diffracted amplitudes $[I_l(R)]^{1/2}$ and $|G_{ll}(R)|$.

4. Compute a modified data set, denoted by $[\hat{I}_l^{obs}(R)]^{1/2}$, as

$$\hat{I}_l^{obs}(R) = I_l^{obs}(R) + K^2 \, |G_{ll}(R)|^2 \, w_{lattice}(R,l/c),$$

where $I_l^{obs}(R)$ is the continuous observed intensity data and K is the scale factor (i.e. $[I_l^{obs}(R)]^{1/2}$ is scaled to $K[I_l^{calc}(R)]^{1/2}$). The modified data $[\hat{I}_l^{obs}(R)]^{1/2}$ represents the continuous data that would be observed if the specimen were perfectly noncrystalline, based on the current estimate of the molecular structure.

5. Refine the molecular structure against $[\hat{I}_l^{obs}(R)]^{1/2}$, treating it as if the specimen were noncrystalline.

6. If step 5 results in no significant change in the x-ray agreement then stop, otherwise return to step 3.

The procedure is repeated for a set of values of σ_{lat} and σ_{axial} to find those that give the best agreement. A model consistent with the continuous x-ray data and incorporating the effect of the disorder is then obtained.

Application to α-poly(dA)·poly(dT)

The polynucleotide duplex poly(dA)·poly(dT) has been trapped in both polycrystalline and screw-disordered fibres. Structure determination based on x-ray data from these different allomorphs shows conservation of the overall molecular morphology, and this system therefore provides an ideal test-bed for the method described above.

Poly(dA)•poly(dT) has been trapped in three distinct forms in oriented fibres. Two of these are polycrystalline, one with one molecule per unit cell (the β_1-form) [10], and the other with two molecules per unit cell (the β_2-form) [11,12]. Structure determination for the two β forms shows that the overall molecular structures are quite similar in the two allomorphs [10,12]. The two strands of the double helix are similar, but are different enough conformationally to give a duplex morphology distinct from that of classical B-DNA. The third form of poly(dA)·poly(dT), denoted the α-form [11], gives a diffraction pattern that contains sharp Bragg reflections at low resolution, that give way to continuous diffraction at higher resolution (Figure 1). Park *et al*. [8] analysed the structure of α-poly(dA)·poly(dT) using the continuous x-ray data between approximately 6 and 3Å resolution, and the low resolution Bragg data. In this analysis, it was assumed that the continuous diffraction at a resolution greater than ~6Å is due to a perfectly noncrystalline specimen, and that the sharp Bragg reflections at a resolution of less than ~6Å are due to a perfectly polycrystalline specimen. This approach was based on the reasonable assumption that the Bessel term $G_{ll}(R)$ which is eliminated by the screw disorder (in the absence of lattice disorder) would have a relatively small amplitude at resolutions greater than 6Å, so that its inclusion in the calculation would introduce only a small error. However, such an approach does involve an approximation, and it also imposes the restriction of having to exclude data with spacings > 6Å. We therefore applied the algorithm described above to the continuous diffraction data from α-poly(dA)·poly(dT).

A survey of different disorder models for α-poly(dA)·poly(dT) by Stroud & Millane [13] confirmed that the diffraction data are best explained by a specimen in which the molecules in the crystallites are randomly screw disordered, and also indicated that the lattice is subject to small lateral and axial distortions. Referring to (4), and considering the behaviour of the Bessel functions and the function $w(R,Z)$, the disorder is expected to affect the diffraction primarily at medium resolution. We therefore conducted two sets of refinements, one using the original diffraction data in the resolution range 6-3Å, and one using data in the region 8-3Å. The medium (8-6Å) resolution continuous data were measured only where there was no interference by Bragg sampling. The structure was refined conventionally against the continuous data using each data set. It was also refined, incorporating the effects of the disorder, by using the algorithm described above, for a range of values of σ_{lat} and σ_{axial}, and using both data sets. The algorithm converged within four cycles in all cases, and the best agreement was obtained for $\sigma_{lat} \approx \sigma_{axial} \approx 0.2$Å. The results of the refinements are listed in Table 1. Referring to the table shows that for the case of data in the range 6-3Å, although the x-ray agreement is slightly better for the refinement based on the screw-disordered model than for that based on the noncrystalline model, the differences are quite small. This is not particularly surprising, and validates the approach taken by Park *et al*. [8], showing that the particular x-ray data they used are indeed relatively insensitive to the disorder as described above. However, the approach is not completely satisfactory since it does not allow one to make use of all the available diffraction data, and in the presence of lattice disorder the term $|G_{ll}(R)|$ does contribute to the continuous diffracted intensity. The results obtained using data in the range 8-3Å show why Park *et al*. [8] had to use the restricted data set. Significantly better agreement is obtained for a refinement based on a screw-disordered specimen than for one based on a noncrystalline specimen (Table 1). With the new algorithm, very good agreement ($R'' = 0.23$) is obtained with many more diffraction data (182 vs. 152) and addition of only one parameter. (We set $\sigma_{lat} = \sigma_{axial}$ since releasing this constraint did not lead to significantly better agreement). As expected, a poor fit with the larger x-ray data set is obtained ($R'' = 0.39$) if the effect of the

Resolution range of data (Å)	Screw disordered	R	R”
6 - 3	N	0.25	0.28
6 - 3	Y	0.22	0.26
8 - 3	N	0.32	0.39
8 - 3	Y	0.20	0.23

Table 1: Conventional (R) and quadratic (R") crystallographic R-factors for the continuous diffraction for various models as described in the text

disorder is not incorporated. The agreement between the measured and calculated amplitudes is shown in Figure 2. The fit of the final model to the Bragg data was assessed by refining only the scale factor while fitting $[I_{il}^{(sd)}]^{1/2}$ to the Bragg data $[I_{il}^{obs}]^{1/2}$. This gave $R = 0.21$ and $R'' = 0.23$ for the Bragg data. The backbone conformation angles of the structure obtained are similar to those of the previously determined a structure, and the β_1 and β_2 structures. Differences between the α structures are generally no larger than the differences amongst the original α structure and the β structures. This is can be seen by comparison of the molecular structure obtained here with that of Ref. 8 (Figure 3).

Discussion

Polymer and other biomolecular structures so far determined by x-ray fibre diffraction analysis have almost always utilized diffraction data from either noncrystalline or polycrystalline specimens. In the few cases where data from specimens with intermediate ordering have been used, the disorder has been only approximately taken into account. Explicit treatment of the effects of the disorder allows use of the full set of available diffraction data, and leads to a more rigorous, accurate and satisfactory structure determination. Application of the method described here for the case of screw-disorder leads to a structure that is within the conformational domain defined by the other crystal forms and gives good x-ray agreement against a larger data set. This lends support to the validity of the approach. For this particular example, the advantages of this approach are relatively small, since the increase in the number of data that it allows is rather modest. However, this will not always be the case. In cases where the disorder affects most, or all, of the diffraction data, using an approach of this kind will be necessary for accurate structure determination.

We are grateful to Drs. R. Chandrasekaran and Struther Arnott for discussion, and the U.S. National Science Foundation for support (DBI-9722862).

References

[1] Millane, R.P., in *Crystallographic Computing 4: Techniques and New Technologies* (eds. Isaacs, N.W. & Taylor, M.R.), Oxford, London, 1987, pp. 169-186.

[2] Millane, R.P., in *International Tables for Crystallography, Vol. B* (ed. Schmueli, U.), Kluwer, Dordrecht, 2001, pp. 466-481.

[3] Miller, A. & Parry, D.A.D., *Polymer* (1974) 15, 706-712.

[4] Arnott, S., in *Fiber Diffraction Methods* (eds. French, A.D. & Gardner, K.H.), American Chemical Society, Washington, DC, 1980, pp. 1-30.

[5] Stroud, W.J. & Millane, R.P., *Acta Cryst.* (1995) A51, 771-790.

[6] Millane, R.P. & Stroud, W.J., *Fiber Diffraction Rev.* (1996) 5, 16-20.

[7] Arnott, S., Chandrasekaran, R., Millane, R.P. & Park, H.S., *J. Mol. Biol*. (1986) 188, 631-640.

[8] Park, H.S., Arnott, S., Chandrasekaran, R. & Millane, R.P., *J. Mol. Biol.* (1987) 197, 513-523.

[9] Smith, P.J.C. & Arnott, S., *Acta Cryst.* (1978) A34, 3-11.

[10] Chandrasekaran, R., Radha, A. & Park, H.S., *Acta Cryst.* (1995) D51, 1025-1035.

[11] Arnott, S. & Selsing, E., *J. Mol. Biol.* (1974) 88, 509-521.

[12] Chandrasekaran, R. & Radha, A., *J. Biomol. Struct. Dynam.* (1992) 10, 153-168.

[13] Stroud, W.J. & Millane, R.P., *Acta Cryst.* (1995) A51, 790-800.